

GREEN'S SUMSET PROBLEM AT DENSITY ONE HALF

TOM SANDERS

ABSTRACT. We investigate the size of subspaces in sumsets and show two main results. First, if $A \subset \mathbb{F}_2^n$ has density at least $1/2 - o(n^{-1/2})$ then $A + A$ contains a subspace of co-dimension 1. Secondly, if $A \subset \mathbb{F}_2^n$ has density at least $1/2 - o(1)$ then $A + A$ contains a subspace of co-dimension $o(n)$.

1. INTRODUCTION

In the paper [Bou90] Bourgain first addressed the question of showing that if $A \subset \{1, \dots, N\}$ has positive relative density then $A + A$ contains a very long arithmetic progression. Since his work the problem has received considerable attention, and to help understand it better Green [Gre05] introduced a model version which has turned out to be interesting in its own right. It is this question with which we shall concern ourselves in this note.

Suppose, as we shall throughout, that $G := \mathbb{F}_2^n$ and let \mathbb{P}_G denote the normalized counting measure on G . We are interested in what size of subspace one can guarantee that $A + A$ contains, where A is a subset of G of density $\alpha := \mathbb{P}_G(A)$.

It turns out that there are various ranges of the density in which we see quite different phenomena. To begin note that if $\alpha > 1/2$ then the inclusion-exclusion principle tells us that $\mathbb{P}_G(A \cap (x + A)) > 0$ for all $x \in G$ and so we have that $A + A = G$; we write this as follows.

Proposition 1.1. *Suppose that $A \subset G := \mathbb{F}_2^n$ has density $\alpha > 1/2$. Then $A + A$ contains a subspace of co-dimension 0.*

Once the density dips below $1/2$ things begin to change. In this regime A may be contained in a subspace of co-dimension 1, and so $A + A$ can, at best, be guaranteed to contain a subspace of co-dimension 1. To start with this is best possible:

Theorem 1.2. *Suppose that $A \subset G := \mathbb{F}_2^n$ has density $\alpha > 1/2 - \epsilon$ where $\epsilon \in (0, 1/2^9 \sqrt{n}]$. Then $A + A$ contains a subspace of co-dimension 1.*

Once $\epsilon \sim 1/\sqrt{n}$, however, a different sort of behavior manifests. The worst known such is exhibited by the so called Niveau set construction of Ruzsa [Ruz87] and for us this yields the following theorem.

Theorem 1.3 (Green-Ruzsa). *For all $\epsilon \in (2^2/\sqrt{n}, 1/2]$, there is a set $A = A(\epsilon) \subset G := \mathbb{F}_2^n$ of density $\alpha > 1/2 - \epsilon$ such that any subspace contained in $A + A$ has co-dimension $\Omega(\epsilon\sqrt{n})$.*

We are not able to prove a matching upper bound, although we are able to establish the following weak complement.

Theorem 1.4. *Suppose that $A \subset G := \mathbb{F}_2^n$ has density $\alpha > 1/2 - \epsilon$ where $\epsilon \in (0, 1/2]$. Then $A + A$ contains a subspace of co-dimension $O(\frac{n}{\log \epsilon^{-1}})$.*

All previous work on showing that sumsets contain large subspaces has concentrated on the case of positive density (rather than density close to $1/2$) and has consequently produced weaker results. It turns out the argument used for the previous theorem also yields a new result in this case.

Theorem 1.5. *Suppose that $A \subset G := \mathbb{F}_2^n$ has density $\alpha > 0$. Then $A + A$ contains a subspace of dimension $\Omega(\alpha n)$.*

This improves upon [Gre05, Theorem 9.3] where a lower bound on the dimension of the form $\Omega(\alpha^2 n)$ was given. Recently Croot and Sisask [CS10] also established an improvement of roughly the strength of Theorem 1.5 but by different arguments in a much more general setting.

Finally, Fourier analysis is notoriously weak when dealing with thin sets and, indeed, if $\alpha = o(\log n/n)$ then it turns out that an elementary counting argument of Croot, Ruzsa and Schoen [CRS07] supersedes Theorem 1.5.

As indicated we make use of the Fourier transform the group $G := \mathbb{F}_2^n$. In particular we denote the dual group by \widehat{G} and define the transform to be the map taking $f \in L^1(G)$ to

$$\widehat{f}(\gamma) := \mathbb{E}_{x \in G} f(x) \overline{\gamma(x)}.$$

We use basic results from Fourier analysis without comment and the reader interested in details may wish to consult Tao and Vu [TV06].

The note now splits into three further sections in which Theorem 1.2, Theorem 1.3 and Theorem 1.4 (and 1.5) are proved, followed by some concluding remarks in the final section including a discussion of the link with the integer version of the problem.

2. PROOF OF THEOREM 1.2

The argument involves two main tools. The first is Plünnecke's inequality, [Plü69], which we record now. One of the key ideas in our work is to make use of it in the region when $K \sim 1$.

Theorem 2.1 ([TV06, Corollary 6.28]). *Suppose that $A, B \subset G := \mathbb{F}_2^n$ are such that $\mathbb{P}_G(A + B) \leq K \mathbb{P}_G(A)$. Then for any positive integer k there is a set $X \subset A$ with $\mathbb{P}_G(X + kB) \leq K^k \mathbb{P}_G(X)$.*

The second tool is measure concentration on the cube for which we shall follow McDiarmid [McD89]. The idea of using measure concentration was introduced to difference set problems by Wolf in [Wol10] and that is the inspiration for our application.

For any natural number n we write Q_n for the cube $\{0, 1\}^n$. The Hamming metric on Q_n is defined in the usual way:

$$d(x, y) := |\{i : x_i \neq y_i\}| \text{ for all } x, y \in Q_n.$$

For any set $A \subset Q_n$ and $r \geq 0$ we write

$$\text{Ham}_r(A) := \{x \in Q_n : d(x, y) \leq r \text{ for some } y \in A\},$$

that is the set of points of at most distance r from A . Measure concentration provides a lower bound for the density of this set.

Theorem 2.2 ([McD89, Proposition 7.7]). *Suppose that A is a non-empty subset of Q_n . Then for any $r \geq 0$ we have*

$$\mathbb{P}_{Q_n}(\text{Ham}_r(A)) \geq 1 - \frac{\exp(-r^2/2n)}{\mathbb{P}_{Q_n}(A)}.$$

The formal similarity of \mathbb{F}_2^n and Q_n immediately tells us how we shall make use of this result. Suppose that $E = \{e_1, \dots, e_n\}$ is a basis of $G := \mathbb{F}_2^n$, so that the map

$$\phi_E : Q_n \rightarrow G; x \mapsto x_1 \cdot e_1 + \dots + x_n \cdot e_n$$

is a bijection. Sets in G inherit certain growth properties from those in Q_n as follows. Writing $F := E \cup \{0_G\}$ we have that

$$(2.1) \quad A + rF \supset \phi_E(\text{Ham}_r(\phi_E^{-1}(A))).$$

To see this note that if z is a member of the right hand side then there is some $y \in A$ such that

$$d(\phi_E^{-1}(z), \phi_E^{-1}(y)) \leq r.$$

Note that this is perfectly well defined since ϕ_E^{-1} is a bijection. Now, let $x \in \{0, 1\}^n$ be such that

$$x_i = \begin{cases} 1 & \text{if } \phi_E^{-1}(z)_i \neq \phi_E^{-1}(y)_i \\ 0 & \text{otherwise,} \end{cases}$$

so that

$$\phi_E^{-1}(z) = \phi_E^{-1}(y + \phi_E(x)).$$

On the other hand the number of i such that $x_i \neq 0$ is at most r and $0_G \in F$, whence $z \in y + rF$. (2.1) then follows. In light of this we have the following consequence of Theorem 2.2.

Proposition 2.3. *Suppose that $G = \mathbb{F}_2^n$ and E is a basis of G , A is a non-empty subset of G and $F := E \cup \{0_G\}$. Then for any $r \geq 0$ we have*

$$\mathbb{P}_G(A + rF) \geq 1 - \frac{\exp(-r^2/2n)}{\mathbb{P}_G(A)}.$$

The heart of the argument is the following asymmetric version of Theorem 1.2.

Proposition 2.4. *Suppose that $A \subset G := \mathbb{F}_2^n$ has density $\alpha > 1/2 - \epsilon$ where $\epsilon \in (0, 1/2^9 \sqrt{n}]$. Then $A + A$ contains a coset of a subspace of co-dimension 1.*

Proof. In view of Proposition 1.1 we may certainly assume that $\mathbb{P}_G(A) \leq 1/2$. Put $S := (A + A)^c$ and note that $(S + A) \cap A = \emptyset$, whence $\mathbb{P}_G(A + S) \leq 1 - \mathbb{P}_G(A)$. Since $\mathbb{P}_G(A) = \alpha > 1/2 - \epsilon$, it follows that

$$\mathbb{P}_G(A + S) < \frac{1 + 2\epsilon}{1 - 2\epsilon} \mathbb{P}_G(A) \leq \exp(6\epsilon) \mathbb{P}_G(A),$$

since $\epsilon \leq 1/4$. Suppose that E is a basis of G such that there is some $s \in S$ with $s + E \subset S$. Since \mathbb{P}_G is translation invariant we have that $F := E \cup \{0_G\}$ has

$$\mathbb{P}_G(A + F) = \mathbb{P}_G(A + (s + F)) \leq \mathbb{P}_G(A + S) \leq \exp(6\epsilon) \mathbb{P}_G(A).$$

Now, let $k = \lceil \sqrt{n} \rceil$ and define a sequence of sets X_0, X_1, \dots using Plünnecke's inequality: let $\emptyset \neq X_0 \subset A$ be such that $\mathbb{P}_G(X_0 + kF) < \exp(6\epsilon k) \mathbb{P}_G(X_0)$, and $\emptyset \neq X_{r+1} \subset X_r$ be such that

$$(2.2) \quad \mathbb{P}_G(X_{r+1} + 2^{r+1}kF) \leq \exp(6\epsilon \cdot 2^{r+1}k) \mathbb{P}_G(X_{r+1}).$$

Since $X_r \neq \emptyset$ and $(X_r)_r$ is nested, we see that the sequence $(\sqrt{2 \log 8 \mathbb{P}_G(X_r)^{-1}})_r$ is increasing and bounded above by $O(\sqrt{n})$. It follows that there is some r such that $2^r \geq \sqrt{2 \log 8 \mathbb{P}_G(X_r)^{-1}}$; let r' be the minimal such. In this case

$$2^{r'-1} < \sqrt{2 \log 8 \mathbb{P}_G(X_{r'-1})^{-1}} \leq \sqrt{2 \log 8 \mathbb{P}_G(X_{r'})^{-1}}$$

by nesting of $(X_r)_r$, whence

$$(2.3) \quad \sqrt{2 \log 8 \mathbb{P}_G(X_{r'})^{-1}} \leq 2^{r'} \leq 2 \cdot \sqrt{2 \log 8 \mathbb{P}_G(X_{r'})^{-1}} \leq 2^5 \log \frac{3 \mathbb{P}_G(X_{r'})^{-1}}{4}.$$

But by Proposition 2.3 we have that

$$\mathbb{P}_G(X_{r'} + 2^{r'}kF) + \frac{\exp(-2^{2r'}k^2/2n)}{\mathbb{P}_G(X_{r'})} \geq 1.$$

Now by (2.2) and the upper bound in (2.3) we get that

$$\mathbb{P}_G(X_{r'} + 2^{r'}kE) \leq \exp(6\epsilon \cdot 2^{r'}k) \mathbb{P}_G(X_{r'}) \leq \frac{3}{4}$$

since $k \leq 2\sqrt{n} \leq 2^{-8}\epsilon^{-1}$ by assumption on ϵ ; whence

$$\frac{\exp(-2^{2r'}k^2/2n)}{\mathbb{P}_G(X_{r'})} \geq \frac{1}{4}.$$

On the other hand this can be bounded above using the lower bound in (2.3) and the fact that $k^2 \geq n$:

$$\frac{\exp(-2^{2r'}k^2/2n)}{\mathbb{P}_G(X_{r'})} \leq \frac{(8 \mathbb{P}_G(X_{r'})^{-1})^{-k^2/n}}{\mathbb{P}_G(X_{r'})} \leq \frac{1}{8}.$$

This contradiction means that for all $s \in S$, the set $S - s$ contains at most $n - 1$ linearly independent vectors. Thus there is an element $s \in S$ and a subspace H of co-dimension 1 in G such that $S \subset s + H$. Since $S = (A + A)^c$ it follows that $A + A \supset (s + H)^c$; $(s + H)^c$ is simply the coset of H not equal to $s + H$, whence we are done. \square

We now use the above result in a straightforward manner.

Proof of Theorem 1.2. Let U be the smallest subspace of G such that $A + A \supset U^c$. Note that such a space exists since $A + A \supset \emptyset = G^c$.

By averaging there is a coset $x + U$ on which A has relative density at least α : write $A' := (A \cap (x + U) - x)$ and note that $A + A \supset A' + A'$ and $\mathbb{P}_U(A') \geq \alpha$. By Proposition 2.4 there is a subspace $U' \leq U$ of relative co-dimension at most 1 and some $u \in U$ such that $u + U' \subset A' + A' \subset A + A$. We have three cases:

(i) $U' = U$: then

$$G = U \cup U^c = U' \cup U^c \subset A + A;$$

(ii) $U' \neq U$ and $u + U' \neq U'$: then $A + A \supset U'^c$ and $\dim U' < \dim U$ contradicting the minimality of U ;

(iii) $U' \neq U$ and $u + U' = U'$: then let $\pi : G \rightarrow U$ be a projection which is the identity when restricted to U and note that $\pi^{-1}(U')$ is a subspace of G of co-dimension 1 with

$$\begin{aligned} \pi^{-1}(U') &= (\pi^{-1}(U') \cap U) \cup (\pi^{-1}(U') \cap U^c) \\ &\subset U' \cup U^c \subset A + A. \end{aligned}$$

The result follows. \square

3. PROOF OF THEOREM 1.3

The argument here is a very slight adaptation of [Gre05, Theorem 9.4]. Green established this by reformulating Ruzsa's construction from [Ruz91] in the model setting where many of the details simplify.

Proof of Theorem 1.3. Let

$$A := \{x \in \mathbb{F}_2^n : x \text{ has at most } n/2 - \eta\sqrt{2\pi n}/2 \text{ ones}\}.$$

Let X be the random variable which takes $x \in G$ to the number of 1s in x . \mathbb{P}_G is the uniform distribution on G , and X is the sum of n independent identically distributed Bernoulli random variables with parameter $p = 1/2$.

The mean of each individual Bernoulli trial is $1/2$ and the variance $1/4$ so that the mean of X is $n/2$ and the variance is $n/4$. It follows from the Berry-Esseen inequality (see, e.g. [Shi96, p374]) that

$$\sup_{x \in G} |\mathbb{P}_G(A) - \Phi(-\eta\sqrt{2\pi})| \leq \frac{3.2}{\sqrt{n}}.$$

Thus

$$\mathbb{P}_G(A) \geq \Phi(-\eta\sqrt{2\pi}) - \frac{3.2}{\sqrt{n}}.$$

On the other hand

$$\Phi(2\lfloor \eta\sqrt{2\pi n}/2 \rfloor / \sqrt{n}) = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^{\eta\sqrt{2\pi}} \exp(-x^2/2) dx \geq \frac{1}{2} - \eta.$$

It follows that

$$\mathbb{P}_G(A) \geq \frac{1}{2} - \eta - \frac{3.2}{\sqrt{n}}.$$

It follows that for $\epsilon\sqrt{n} \geq 2^2$ we may pick $\eta = 0.8\epsilon$ and get that $\mathbb{P}_G(A) > 1/2 - \epsilon$.

Now we shall show that if $V \leq G$ has co-dimension at most $d := \lfloor \eta\sqrt{2\pi n} \rfloor$ then V contains a vector with at most $\lfloor \eta\sqrt{2\pi n} \rfloor$ zeros in the standard basis. Since any $x \in A + A$ has at least $\eta\sqrt{2\pi n}$ zeros in the standard basis we shall be done. Such a V can be written as

$$V = \{\lambda_1 v_1 + \cdots + \lambda_{n-d} v_{n-d} : \lambda_i \in \mathbb{F}_2\},$$

where the v_i s are linearly independent. The v_i s may be written in the standard basis as

$$v_i = \epsilon_i^{(1)} e_1 + \cdots + \epsilon_i^{(n)} e_n$$

where $(e_i)_i$ is the standard basis. The column rank of the matrix $(\epsilon_i^{(j)})_{ij}$ is $n - d$ hence so is its row rank. Without loss of generality we may suppose that the first $n - d$ rows $(\epsilon_1^{(j)}, \dots, \epsilon_{n-d}^{(j)})$, $j = 1, \dots, n - d$ are linearly independent. It follows that we can solve the $n - d$ equations

$$\lambda_1 \epsilon_1^{(j)} + \cdots + \lambda_{n-d} \epsilon_{n-d}^{(j)} = 1$$

for the λ_i giving a vector in V with no more than d zeros. The result follows. \square

4. PROOF OF THEOREM 1.4

We shall prove the following stronger theorem from which both Theorems 1.4 and 1.5 follow.

Theorem 4.1. *Suppose that $A \subset G := \mathbb{F}_2^n$ has density $\alpha \leq 1/2$. Then $A + A$ contains a subspace of co-dimension*

$$\left\lceil n / \log_2 \frac{2 - 2\alpha}{1 - 2\alpha} \right\rceil.$$

Before proving this we establish our two consequences.

Proof of Theorem 1.4. Apply Theorem 4.1 with $\alpha = 1/2 - \epsilon$. \square

Proof of Theorem 1.5. It is easy to see that

$$\left\lceil n / \log_2 \frac{2 - 2\alpha}{1 - 2\alpha} \right\rceil \leq n(1 - \alpha / \log 2) + O(1).$$

Theorem 4.1 then tells us that $A + A$ contains a subspace of dimension at least $\alpha n / \log 2 - O(1)$. \square

The proof is inspired by the standard iterative method of Roth introduced in [Rot52] and the more commonly cited [Rot53], which was adapted to finite fields by Meshulam in [Mes95]. The following lemma is the driver.

Lemma 4.2 (Iteration lemma). *Suppose that $A \subset G := \mathbb{F}_2^n$ has density $\alpha > 0$ and $V \leq G$. Then there is a subspace $V' \leq V$ of relative co-dimension 1 such that*

$$\mathbb{P}_{V'}(V' \setminus (A + A)) \leq \frac{1 - 2\alpha}{1 - \alpha} \mathbb{P}_V(V \setminus (A + A)).$$

Proof. For each $W \in G/V$ fix some $x_W \in W$ and let A_W be $A \cap W - x_W$ considered as a subset of V . Since $-2x_W = 0$ we have that $A_W + A_W = A \cap W + A \cap W$, whence

$$(A + A) \cap V = \bigcup_{W \in G/V} (A_W + A_W);$$

write $S := V \setminus (A + A)$. In view of the definition of S and our above observation, we get from Plancherel's theorem that

$$(4.1) \quad 0 = \langle \sum_{W \in G/V} 1_{A_W} * 1_{A_W}, 1_S \rangle_{L^2(V)} = \sum_{\gamma \in \widehat{V}} \sum_{W \in G/V} |\widehat{1_{A_W}}(\gamma)|^2 \widehat{1_S}(\gamma).$$

Partition \widehat{V} into two sets $\mathcal{N} := \{\gamma \in \widehat{V} : \widehat{1_S}(\gamma) < 0\}$ and $\mathcal{P} := \{\gamma \in \widehat{V} : \widehat{1_S}(\gamma) \geq 0\}$. Since $0_{\widehat{V}} \in \mathcal{P}$ we have

$$\sum_{\gamma \in \mathcal{P}} \sum_{W \in G/V} |\widehat{1_{A_W}}(\gamma)|^2 \widehat{1_S}(\gamma) \geq \mathbb{P}_V(S) \sum_{W \in G/V} \mathbb{P}_V(A_W)^2.$$

On the other hand, from equation (4.1) and the definition of \mathcal{N} , we get

$$\begin{aligned} \sum_{\gamma \in \mathcal{N}} \sum_{W \in G/V} |\widehat{1_{A_W}}(\gamma)|^2 |\widehat{1_S}(\gamma)| &= \sum_{\gamma \in \mathcal{N}} \sum_{W \in G/V} -|\widehat{1_{A_W}}(\gamma)|^2 \widehat{1_S}(\gamma) \\ &= \sum_{\gamma \in \mathcal{P}} \sum_{W \in G/V} |\widehat{1_{A_W}}(\gamma)|^2 \widehat{1_S}(\gamma) \\ &\geq \mathbb{P}_V(S) \sum_{W \in G/V} \mathbb{P}_V(A_W)^2. \end{aligned}$$

By the Hölder's inequality and then Parseval's theorem we have

$$\begin{aligned} \sum_{\gamma \in \mathcal{N}} \sum_{W \in G/V} |\widehat{1_{A_W}}(\gamma)|^2 |\widehat{1_S}(\gamma)| &\leq \sup_{\gamma \in \mathcal{N}} |\widehat{1_S}(\gamma)| \sum_{\gamma \neq 0_{\widehat{G}}} \sum_{W \in G/V} |\widehat{1_{A_W}}(\gamma)|^2 \\ &= \sup_{\gamma \in \mathcal{N}} |\widehat{1_S}(\gamma)| \sum_{W \in G/V} (\mathbb{P}_V(A_W) - \mathbb{P}_V(A_W)^2). \end{aligned}$$

Combining this with the foregoing we get that

$$(4.2) \quad \mathbb{P}_V(S) \sum_{W \in G/V} \mathbb{P}_V(A_W)^2 \leq \sup_{\gamma \in \mathcal{N}} |\widehat{1_S}(\gamma)| \sum_{W \in G/V} (\mathbb{P}_V(A_W) - \mathbb{P}_V(A_W)^2).$$

Now, by the Cauchy-Schwarz inequality

$$\mathbb{E}_{W \in G/V} \mathbb{P}_V(A_W) \cdot \sum_{W \in G/V} \mathbb{P}_V(A_W) \leq \sum_{W \in G/V} \mathbb{P}_V(A_W)^2.$$

However, $\mathbb{E}_{W \in G/V} \mathbb{P}_V(A_W) = \alpha$, whence

$$\sum_{W \in G/V} (\mathbb{P}_V(A_W) - \alpha)^2 \leq (\alpha^{-1} - 1) \sum_{W \in G/V} \mathbb{P}_V(A_W)^2.$$

Inserting this into (4.2) we get that

$$\mathbb{P}_V(S) \sum_{W \in G/V} \mathbb{P}_V(A_W)^2 \leq \sup_{\gamma \in \mathcal{N}} |\widehat{1}_S(\gamma)| (\alpha^{-1} - 1) \sum_{W \in G/V} \mathbb{P}_V(A_W)^2.$$

Since $\alpha > 0$ and G/V is finite we may divide by the sum and conclude that

$$\sup_{\gamma \in \mathcal{N}} |\widehat{1}_S(\gamma)| \geq \frac{\alpha}{1 - \alpha} \mathbb{P}_V(S).$$

Let $\gamma \in \mathcal{N}$ be such that this supremum is attained and write $V' := \{\gamma\}^\perp$. V' has relative co-dimension 1 and, in view of this, we note that $V \setminus V' = x_0 + V'$ for any $x_0 \in V \setminus V'$. Now

$$\mathbb{P}_V(S \cap V') - \mathbb{P}_V(S \cap (x_0 + V')) = \widehat{1}_S(\gamma) \leq -\frac{\alpha}{1 - \alpha} \mathbb{P}_V(S),$$

since $V' = \{\gamma\}^\perp$ and $\gamma \in \mathcal{N}$. Furthermore, since V' and $x_0 + V'$ partition V we have that

$$\mathbb{P}_V(S \cap V') + \mathbb{P}_V(S \cap (x_0 + V')) = \mathbb{P}_V(S).$$

Adding these two expressions tells us that

$$\mathbb{P}_{V'}(S \cap V') = 2 \mathbb{P}_V(S \cap V') = \left(1 - \frac{\alpha}{1 - \alpha}\right) \mathbb{P}_V(S) = \frac{1 - 2\alpha}{1 - \alpha} \mathbb{P}_V(S).$$

It remains only to note that $S \cap V' = V' \setminus (A + A)$ and the lemma is proved. \square

Proof of Theorem 4.1. Let $\sigma \in (0, 1)$ be a parameter to be optimized later. We use the iteration lemma to produce a sequence of subspaces V_i such that

- (i) $V_i \leq V_{i-1}$ and the co-dimension of V_i in V_{i-1} is 1;
- (ii) $\mathbb{P}_{V_i}(V_i \setminus (A + A)) \leq \frac{1-2\alpha}{1-\alpha} \mathbb{P}_{V_{i-1}}(V_{i-1} \setminus (A + A))$.

We set $V_0 = G$, and apply the iteration lemma (Lemma 4.2) repeatedly to get the sequence. Now, if $|V_i \setminus (A + A)| < 1$ then $V_i \setminus (A + A) = \emptyset$, whence $A + A$ contains a subspace of co-dimension i by (i). In view of (i) and (ii) this certainly happens if

$$|V_i| \mathbb{P}_{V_i}(V_i \setminus (A + A)) \leq 2^{n-i} \left(\frac{1-2\alpha}{1-\alpha}\right)^i < 1;$$

taking i minimal such that this inequality is satisfied yields the result. \square

5. CONCLUDING REMARKS

As noted in the introduction Theorem 1.2 is best possible, however there is still a large gap between Theorem 1.3 and Theorem 1.4. It has been suggested in [Gre05] that the truth is closer to Theorem 1.3 in the case when the density of the set is $\Omega(1)$. One might take Theorem 1.2 as some support of this conjecture (at least in the case of density $1/2 - o(1)$). An intermediate question might be the following.

Question 5.1. Suppose that $A \subset G := \mathbb{F}_2^n$ has density $\alpha > 1/2 - C/\sqrt{n}$. Does $A + A$ contain a subspace of co-dimension $O_C(1)$?

Theorem 1.3 tells us that this cannot be sublinear, and while showing that may be hard it could be that $O(C^2)$ is rather more accessible.

Theorem 1.4 (and Theorem 1.2) are the first results which provide a sensible upper bound on the co-dimension rather than lower bound on the dimension of the subspace found in $A + A$, and given the proof one might imagine an improvement to Theorem 1.4 would be possible.

The following is a well-known theorem (see, for example, Metsch [Met03]).

Theorem 5.2. *Suppose that $G := \mathbb{F}_2^n$ and $S \subset G \setminus \{0_G\}$ meets every subspace of dimension d . Then $|S| \geq 2^{n+1-d} - 1$.*

This result can be used in the proof of Theorem 4.1 above to tell us that once we have

$$|V_i \setminus (A + A)| < 2^{n-i+1-d}$$

in the iteration then $A + A$ must contain a subspace of dimension d . This certainly happens when

$$2^{1-d} > \left(\frac{1-2\alpha}{1-\alpha} \right)^i;$$

again taking i minimal such that this inequality holds we get a saving of 1 in the co-dimension. This is, of course, not particularly impressive and, indeed, no naïve attack along these lines will work as it turns out that Theorem 5.2 is also best possible. (Again, see Metsch [Met03].) This may be contrasted with the following result of Alon.

Theorem 5.3 ([Alo07, Theorem 4.1]). *Suppose that $G := \mathbb{F}_2^n$ and $S \subset G \setminus \{0_G\}$ is such that $|S| \leq c\sqrt{|G|/\log |G|}$. Then there is a set $A \subset G$ such that $S = (A + A)^c$.*

Here, of course, the set A produced is rather thin and certainly nowhere near the densities we are looking for.

One may also reasonably ask what happens in the transition to the integers. Theorem 1.5 can be proved there through the machinery of Bohr sets as developed by Bourgain [Bou99]. However, as the reader will have realised from the proof, the strength comes from the rather precise subgroup structure which is not present in general and as a result the conclusions are weaker than what is already known [Gre02].

ACKNOWLEDGEMENTS

The author should like to thank Endre Szemerédi and Julia Wolf for useful conversations.

REFERENCES

- [Alo07] N. Alon. Large sets in finite fields are sumsets. *J. Number Theory*, 126(1):110–118, 2007.
- [Bou90] J. Bourgain. On arithmetic progressions in sums of sets of integers. In *A tribute to Paul Erdős*, pages 105–109. Cambridge Univ. Press, Cambridge, 1990.
- [Bou99] J. Bourgain. On triples in arithmetic progression. *Geom. Funct. Anal.*, 9(5):968–984, 1999.
- [CRS07] E. S. Croot, I. Z. Ruzsa, and T. Schoen. Arithmetic progressions in sparse sumsets. In *Combinatorial number theory*, pages 157–164. de Gruyter, Berlin, 2007.
- [CS10] E. S. Croot and O. Sisask. A probabilistic technique for finding almost-periods of convolutions. *Geom. Funct. Anal.*, 2010. To appear.
- [Gre02] B. J. Green. Arithmetic progressions in sumsets. *Geom. Funct. Anal.*, 12(3):584–597, 2002.
- [Gre05] B. J. Green. Finite field models in additive combinatorics. In *Surveys in combinatorics 2005*, volume 327 of *London Math. Soc. Lecture Note Ser.*, pages 1–27. Cambridge Univ. Press, Cambridge, 2005.
- [McD89] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.
- [Mes95] R. Meshulam. On subsets of finite abelian groups with no 3-term arithmetic progressions. *J. Combin. Theory Ser. A*, 71(1):168–172, 1995.
- [Met03] K. Metsch. Blocking sets in projective spaces and polar spaces. *J. Geom.*, 76(1-2):216–232, 2003. Combinatorics, 2002 (Maratea).
- [Plü69] H. Plünnecke. *Eigenschaften und Abschätzungen von Wirkungsfunktionen*. BMwF-GMD-22. Gesellschaft für Mathematik und Datenverarbeitung, Bonn, 1969.
- [Rot52] K. F. Roth. Sur quelques ensembles d’entiers. *C. R. Acad. Sci. Paris*, 234:388–390, 1952.
- [Rot53] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.
- [Ruz87] I. Z. Ruzsa. Essential components. *Proc. London Math. Soc.* (3), 54(1):38–56, 1987.
- [Ruz91] I. Z. Ruzsa. Arithmetic progressions in sumsets. *Acta Arith.*, 60(2):191–202, 1991.
- [Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1996. Translated from the first (1980) Russian edition by R. P. Boas.
- [TV06] T. C. Tao and H. V. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.
- [Wol10] J. Wolf. The structure of popular difference sets. *Israel J. Math.*, 179:253–278, 2010.

DEPARTMENT OF PURE MATHEMATICS AND MATHEMATICAL STATISTICS, UNIVERSITY OF CAMBRIDGE, WILBERFORCE ROAD, CAMBRIDGE CB3 0WA, ENGLAND

E-mail address: t.sanders@dpmps.cam.ac.uk